

When invasion increases population genetic structure: a study with *Centaurea diffusa*

Robin A. Marrs · René Sforza · Ruth A. Hufbauer

Received: 10 July 2007 / Accepted: 26 July 2007 / Published online: 15 August 2007
© Springer Science+Business Media B.V. 2007

Abstract Biological invasions offer excellent systems to study the evolutionary processes involved in introductions of species to new ranges. Molecular markers can reveal invasion histories and the effects of introductions on amounts and structuring of genetic variation. We used five polymorphic microsatellite loci to elucidate genetic diversity and population structure between native range and introduced range populations of a prominent North American rangeland weed, *Centaurea diffusa* (Asteraceae). We found that the total number of alleles and the number of private alleles was slightly higher in the native Eurasian range, and that allelic richness did not differ between the ranges, indicating overall levels of diversity were similar in Eurasia and North America. It therefore seems unlikely that this invasion has been affected by genetic bottlenecks or founder effects. Indeed, results of assignment tests suggest that multiple introductions have contributed

to North America's *C. diffusa* invasion. Additionally, assignment tests show that both Eurasian and North American sites had a strong pattern of mixed genetic ancestry. This mixed assignment corresponded to a lack of geographic population structure among Eurasian samples. The lack of population structure in the native range conflicts with general expectations and findings to date for invasion genetics, and cautions that even species' native ranges may show signs of recent ecological upheaval. Despite the mixed assignments, North American samples showed strong population structure, suggesting that the invasion has been characterized by long-range dispersal of genetically distinct propagules across the introduced range.

Keywords *Centaurea diffusa* · *Centaurea maculosa* · *Centaurea stoebe* · Diffuse knapweed · Genetic diversity · Invasive species · Microsatellites · Population genetic structure

R. A. Marrs · R. A. Hufbauer
Department of Bioagricultural Sciences and Pest
Management, Colorado State University, Fort Collins,
CO 80523, USA

R. A. Marrs (✉) · R. A. Hufbauer
Graduate Degree Program in Ecology, Colorado State
University, Fort Collins, CO 80523, USA
e-mail: robin.mcginn@gmail.com

R. Sforza
European Biological Control Laboratory, USDA-ARS,
Montpellier, France

Introduction

Human-mediated transport of species into novel ecosystems can substantially alter the genetic variation of introduced populations relative to native populations, both in total amount and in how variation is partitioned within and between populations. It is becoming widely recognized that adaptive evolution of introduced species to novel ecosystems

(e.g., Blossey and Notzold 1995; Bosssdorf et al. 2005) plays an important role in biological invasions. Because genetic variation provides the raw materials for adaptive evolution, it is critical to understand how introductions affect the amount and structure of genetic variation. Genetic variation can be measured most easily by using neutral molecular markers, though it is quantitative genetic variation in phenotypic traits that is required for adaptive evolution. A pair of recent meta-analyses explored the relationship between neutral molecular marker diversity and quantitative trait variation; Merila and Crnokrak (2001) found a strong positive correlation, while Reed and Frankham (2001) found a weak positive correlation. Thus, variation in molecular markers appears to be an adequate proxy for overall genetic variation. Molecular genetic data are often used to draw conclusions about how genetic diversity has changed following an introduction (see Durka et al. 2005; Genton et al. 2005; Grapputo et al. 2005 for current examples). Additionally, the insights gleaned by studying the population genetics of invasive species can provide evidence for the invasion patterns of introduced populations (e.g., of multiple introductions, Durka et al. 2005, or of single founding events, Grapputo et al. 2005). Such information can guide research into control efforts. For example, Goolsby et al. (2004) inferred the origin of Old World climbing fern invasive to Florida, USA and used the information to find a potential biological control agent that is highly host specific to the invasive genotypes of the fern from that area of the native range.

When a species is introduced to a new range, the genetic variation within the newly founded populations can be reduced, increased, or unchanged relative to that in the native range. Reduction of genetic diversity is a frequent outcome of species introductions (Amsellam et al. 2000; DeWalt and Hamrick 2004; Grapputo et al. 2005; Bosssdorf et al. 2005). Founder effects and bottlenecks in population size have long been thought to be the dominant processes influencing genetic diversity in species' new ranges (Husband and Barrett 1991). These processes reduce genetic variability in the invaded range as a result of sampling effects or reduced population size (Nei et al. 1975). Selection can also lead to a pattern of reduced genetic diversity in the introduced range of a species. When one introduced genotype is more fit in the new range, its numbers will

tend to increase relative to other genotypes, leading to a reduction in genetic variability (Kliber and Eckert 2005). If these types of processes predominate in an introduction, the introduced populations should be genetically depauperate when compared to native populations (Amsellam et al. 2000; DeWalt and Hamrick 2004; Grapputo et al. 2005).

Alternatively, genetic diversity can be higher within introduced populations than within native populations if introduced populations are founded from multiple, genetically differentiated regions within the native range. Crossing of genetically distinct lineages can lead to increased heterozygosity or novel combinations of alleles (Ellstrand and Schierenbeck 2000). Several recent studies have shown patterns consistent with multiple introductions and high genetic diversity in invaders (Neuffer and Hurka 1999; Kolbe et al. 2004; Williams et al. 2005; Durka et al. 2005; Genton et al. 2005). Because increased genetic diversity also increases the raw materials necessary for adaptive evolution, an invader with high diversity may be quicker to evolve in its new range. A final alternate evolutionary outcome of a species introduction is that of no appreciable change in genetic diversity between the native and introduced ranges. This is a possibility if large population sizes are maintained during (Brown and Marshall 1981) or regained shortly after introductions (Zenger et al. 2003).

The way genetic variation is partitioned within and among populations, or the population structure, can differ between species' original and new ranges. In general, population structure develops when populations differentiate genetically either by adapting to local conditions or via random genetic drift. While local adaptation has been shown in some cases to be quite rapid (e.g., metal tolerance in plants, Pauwels et al. 2005), in general the differentiation process takes time, particularly in neutral markers that evolve by genetic drift. Population structure can be apparent at small geographic scales if a species has a relatively short dispersal distance and individuals are likely to mate with others that are spatially close to them.

As pointed out by Roderick and Navajas (2003), it is often assumed that population structure will be lower in an organisms' introduced range than in its native range. This assumption has been borne out by multiple sets of data on plant introductions (reviewed in Bosssdorf et al. 2005). Three processes may underlie

this pattern. First, for recent introductions, there may not have been enough time for local adaptation or drift to lead to differentiation of populations in the new range. Second, the introduction and subsequent spread in a novel range represent major dispersal events and are likely linked to high gene flow, which minimizes population structure. Finally, if an introduction has imposed a bottleneck in population size that resulted in reduced genetic diversity, it will take time for new variation to arise and population structure to develop, even with a slow rate of spread. It is conceivable, however, that a pattern of lower population structure in the native range than in the introduced range could be observed if the native populations experienced high rates of gene flow, perhaps mediated by humans, and/or the invasive populations were founded by multiple introductions of distinct genotypes into geographically separate areas of the new range. This pattern could be propagated throughout the introduced range by subsequent discrete founding events, leading to a pattern of population structure without isolation by distance.

Here, we examine genetic diversity and population structure in the native and introduced ranges of a prominent rangeland weed, *Centaurea diffusa* Lam. (Asteraceae) using microsatellite loci. Our questions are: (1) Is genetic diversity in the introduced range of *C. diffusa* reduced or increased within populations relative to the native range? (2) Is genetic variation structured similarly between the two ranges? If so, is that structure organized in accordance with the expectations of isolation by distance? (3) Are multiple introductions likely to have occurred in this system? Do the Eurasian samples represent possible areas of origin of our invasive populations?

Methods

Study species

Centaurea diffusa Lam. (Asteraceae) is an annual to triennial species, with the length of its life cycle dependent on resource availability (Thompson and Stout 1991). It is an obligate outcrosser that adopts a tumbling habit for seed dispersal (Watson and Renney 1974). Native range *C. diffusa* (Eurasia) comprises both diploid ($2x = 2n = 18$) and tetraploid ($2x = 4n = 36$) cytotypes. North American *C. diffusa* appear to be diploid (A. Blair, personal

communication). The first *C. diffusa* specimen collected in North America was found in an alfalfa field in Bingen, Washington in 1907 (Watson and Renney 1974). One hundred years later, the species is distributed widely across North America, especially in xeric rangelands of the western USA (LeJeune and Seastedt 2001). The ecological destruction caused by this plant in its invaded range has spurred intensive efforts at biological control; 13 insect species have been introduced to North America as natural enemies of *C. diffusa* (Müller-Scharer and Schroeder 1993; Lang et al. 2000).

Sample collection and preparation

We sampled *C. diffusa* from six Eurasian and eight North American locations (Table 1). Eurasian sampling sites were clustered around the Black Sea (Fig. 1), within the suspected areas of origin of the invasive populations, but were not able to include areas further east (e.g., Turkmenistan), also suspected as a source of the introduction via contaminated alfalfa seed (Roché Jr and Roché 1999). North American sites were spread over much of the invasive range of the species (Fig. 1). Sampling locations with less than 30 individuals were sampled exhaustively, while 30 or more plants were sampled in locations containing many individuals. When a site was not sampled exhaustively, samples were taken at least 1 m apart to reduce the chance of sampling siblings and to capture the genetic variation present at the site. Depending upon the season of collection, either leaf tissue or mature seedheads were collected from individual plants. Leaf tissue was stored on desiccant for transport to the laboratory, while seeds were kept separate by maternal plant, then germinated to provide fresh leaf tissue for DNA extraction. Genomic DNA was extracted from desiccated and fresh leaves using Qiagen Plant-Mini kits (Qiagen, Valencia, CA, USA) and stored at -20°C until genotyping.

Microsatellite analysis

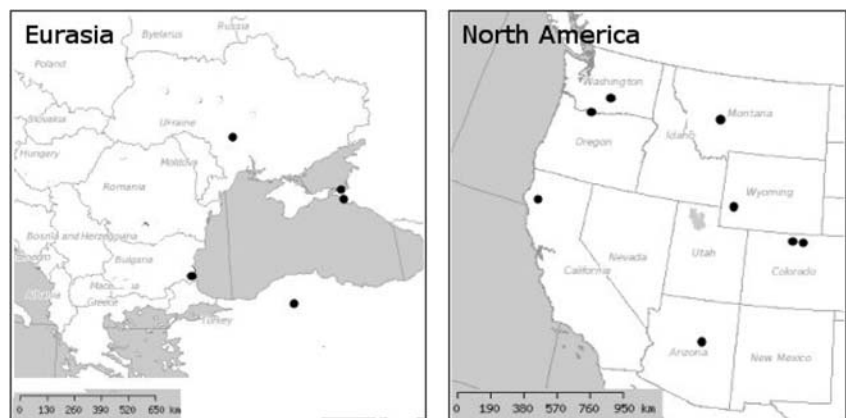
We genotyped the DNA samples using five microsatellite loci (21CM36, 42CM27, 38CM22, CM15, and CM26 from Marrs et al. 2006). Polymerase chain

Table 1 Sampling locations of *C. diffusa*, site codes used in Table 3 and Fig. 3, sample size (N), approximate GPS coordinates, percent polymorphic loci (P), average number of alleles per locus (A_A), number of private alleles (A_P : sampling location A_P values are the number of alleles unique to a single

sampling location; North America and Eurasia A_P values are the number of alleles unique to the continent, including alleles that appeared in more than one sampling location), and gene diversity (H) with standard deviation (SD)

Sampling location	Site code	N	GPS coordinates	P	A_A	A_P	H	SD
North America						16		
Afton, WY	USWY1	9	42.12N 110.15W	100	3.8	0	0.4917	0.3141
Helena, MT	USMT1	25	46.63N 112.00W	100	6.4	2	0.5913	0.3493
Bingen, WA	USWA1	34	45.72N 121.50W	100	6.8	1	0.5081	0.3066
Kittitas, WA	USWA2	14	46.62N 120.37W	100	6.2	0	0.6917	0.4063
Trinity, CA	USCA2	10	41.01N 121.95W	100	4.8	2	0.4450	0.2846
Fort Collins, CO	USCO1	29	40.57N 105.22W	100	5.4	1	0.4865	0.2976
Red Feather, CO	USCO2	27	39.60N 105.00W	100	6.6	2	0.6197	0.3627
Sedona, AZ	USAZ1	14	34.83N 111.15W	100	4.6	1	0.4365	0.2773
Eurasia						28		
Panitsovo, Bulgaria	BG34	20	42.36N 27.71E	100	7.4	5	0.5559	0.3336
Taiman Bay, Russia	RUS5	28	45.33N 37.22E	100	8.0	3	0.5904	0.3481
Anapa, Russia	RUS15	14	44.94N 37.32E	100	5.2	2	0.3594	0.2432
Ürgüp, Turkey	TR6	15	38.69N 34.87E	80	3.8	5	0.3107	0.2271
Cankiri, Turkey	TR10	2	40.76N 33.53E	n/a	n/a	1	n/a	n/a
Pervomajsk, Ukraine	UA3	23	48.09N 30.75E	100	6.8	3	0.5924	0.3508

Fig. 1 Locations of sampled sites in Eurasia and North America



reaction, allele visualization, and band scoring were performed according to Marrs et al. (2006). Multilocus microsatellite genotype data were analyzed by several different methods. To evaluate assumptions of neutrality by determining if any sampling sites deviated from the expectations of Hardy–Weinberg equilibrium and to test for linkage disequilibrium, we used the program GENEPOP on the web (<http://wbiomed.curtin.edu.au/genepop/index.html>,

Raymond and Rousset 1995). We calculated several measures of genetic diversity, including the number of alleles found in each sampling location, the number of private alleles (Neel 1973; Slatkin 1985) in each sampling location, and allelic richness (number of alleles per population corrected for sample size with rarefaction) in each range of the species using FSTAT Version 2.9.3 (Goudet 2001). We calculated gene diversity as expected

heterozygosity with standard deviation for each sampling location using Arlequin Version 3.01 (Excoffier et al. 2005). We used analysis of covariance to compare private alleles per sampling site, total alleles per sampling site, and expected heterozygosity between European and North American *C. diffusa* with continent as the predictor variable and sample size of each population as the covariate (JMP Version 5.0, SAS Institute). Using the numbers of private alleles found per population, we calculated estimates of gene flow (migration) over all sampling sites, in North America, and in Eurasia (Slatkin 1985). Population genetic structuring in the native and introduced ranges were compared with an analysis of molecular variance (AMOVA) computed in Arlequin (Excoffier et al. 2005). Three sets of groups were compared with AMOVA: among continents (North America and Eurasia), among populations within Eurasia, and among populations within North America. In each case permutations to determine significance were set at 10,000 and the distance method used to determine variation in the dataset was the number of different alleles present in each group. As another way to look at population differentiation, we computed pairwise F_{ST} (Weir and Cockerham 1984) with P -values for each pair of sampling locations, excluding Cankiri, Turkey with $N = 2$, using Arlequin (Excoffier et al. 2005). We used a conservative Bonferroni correction to adjust α to equal $P \leq 0.05$ over all comparisons when determining significance. To test whether genetic distance was correlated to geographical distance (isolation by distance) we plotted pairwise F_{ST} against the geographic distance separating sampling locations. Then we calculated Slatkin's (1993) similarity measure ($M = ((1/F_{ST}) - 1)/4$) for each pairwise comparison and used the program IBD (Bohonak 2002) to determine if a significant relationship exists between genetic and geographic distances. Finally, we used the assignment test software Structure Version 2.0 (Pritchard et al. 2000; Falush et al. 2003) to infer the number of genetic clusters present in the dataset and to assign individuals to these clusters. The model uses a Bayesian approach. Model-based methods like Bayesian inference fundamentally assume that observations from each cluster are randomly drawn and that all potential source populations are predefined (Manel et al. 2005). Structure also assumes that Hardy–Weinberg equilibrium is present within

populations and that there is complete linkage equilibrium between loci within populations (Pritchard et al. 2000). We used Pritchard et al.'s (2000) ad hoc method for determining the approximate number of genetic clusters (K = number of clusters of individuals characterized by a set of allele frequencies at each locus) by giving the program a range of values for K as priors and determining which one gave the highest estimated log probability of the data. We computed three independent runs for each possible K from two to 18. The three runs were consistent for all values of K and assured that a burn-in of 250,000 followed by 750,000 data collection repetitions was sufficient to reach a stable α and estimated log probability of the data. We selected the admixture model, which assumes that individuals may have mixed ancestry (each individual is modeled as having inherited some fraction of its genome from ancestors) because of the likelihood of interpopulation and interspecific crossing in the *C. diffusa* system. We also selected to model allele frequencies as independent between populations, a prior that expects allele frequencies in different populations to be somewhat different from one another (Falush et al. 2003).

Results

Hardy–Weinberg and linkage equilibria

Probability tests showed that the majority of the locus by sampling location comparisons did not deviate significantly from Hardy–Weinberg equilibrium. When corrected for multiple comparisons using the Bonferroni method, 13 of 70 comparisons deviated from expectations and showed a deficit of heterozygotes. Of these 13 deviations, seven occurred at locus 38CM22. The remainder of the heterozygote deficiencies occurred without pattern at the other loci. Only three sampling locations had more than one locus out of Hardy–Weinberg equilibrium: in Eurasia, the Pervomajsk, Ukraine site had three loci significantly deficient in heterozygotes, and Taiman Bay, Russia had two loci that deviated in the same manner, while in North America Helena, MT had two loci that showed heterozygote deficiency. Linkage disequilibrium was not significant at any locus by locus comparison.

Genetic diversity

All loci were polymorphic in all sampling locations, except for locus 21CM36 in Ürgüp, Turkey (Table 1). Overall, 76 of the 92 total recorded alleles appeared in Eurasia, and 64 were observed in North America. The average number of alleles observed per locus ranged from 3.8 to 8.0 for Eurasian sites, and from 3.8 to 6.8 for North American sites (Table 1). There was no significant effect of region on the number of alleles found per sampling site (EU mean = 6.24, NA mean = 5.58, $F_{1,12} = 0.89$, $P = 0.368$), though sample size significantly affected the number of alleles observed at a site ($F_{1,12} = 17.13$, $P = 0.002$). We found 28 alleles private (exclusive) to Eurasia; nine of these were found in more than one Eurasian sampling location, and 19 were unique to individual sites. North America contained 16 total private alleles, nine of which were unique to a single sampling location and seven of which were found at more than one North American site. The number of private alleles observed per sampling site was significantly higher in Eurasia than in North America (EU mean = 3.60, NA mean = 1.13, $F_{1,12} = 16.00$, $P = 0.003$). Sample size did not have a significant effect on the number of private alleles per site ($F_{1,12} = 0.08$, $P = 0.390$). Each sampling location was found to have alleles unique to it except two of the North American sites, Afton, WY and Kittitas, WA. In spite of the greater number of private alleles present in the native range of *C. diffusa*, allelic richness in the Eurasian ($R_S = 1.649$) and

North American ($R_S = 1.638$) regions did not differ significantly ($P = 0.883$). Expected heterozygosity was slightly higher in North America than in Eurasia, but the difference was not significant (EU mean = 0.482, NA mean = 0.534, $F_{1,12} = 2.76$, $P = 0.131$). Population sample size had a significant effect on expected heterozygosity ($P = 0.014$), and there was a significant interaction between population sample size and region ($P = 0.049$).

Population structuring

Analysis of molecular variation (AMOVA, Table 2) showed a high degree of within-population variation. In each AMOVA the within-population variance component accounted for at least 90% of the variation. In the continental comparison (native + introduced ranges), none of the variation was due to continent-level differences, while among-population variation within continents was significant. The Eurasian comparison showed a barely significant among-population variance component explaining 2.93% of the variance, while the within-population component was highly significant and explained the remainder of the variance. North American sampling locations showed a stronger pattern of population structuring, with 10.38% of the variance explained by the among-population component. Even here, most of the variance was still explained by within-population variation.

Table 2 Analysis of molecular variance (AMOVA) results

Source of variation	df	Sum of squares	Variance components	Percentage of variation
Native + introduced ranges				
Between continents	1	11.099	0.02061	1.41(n/s)
Among sites, within continents	12	63.353	0.11503	7.86****
Within sites	486	645.13	1.32743	90.73****
Native range (Europe)				
Among sites	5	12.056	0.03876	2.93*
Within sites	180	231.38	1.28544	97.07****
Introduced range (North America)				
Among sites	7	51.298	0.15664	10.38****
Within sites	306	413.75	1.35212	89.62****

Distance method used was R_{ST} -like (sum of squared differences). 10,000 permutations were run to determine significance

* $P < 0.05$

**** $P < 0.00001$

Pairwise F_{ST} comparisons (Table 3) also demonstrated an overall trend of more population-level structuring in North America than in Eurasia. Only three of the ten possible Eurasian to Eurasian comparisons had a significant F_{ST} ($P \leq 0.05$, adjusted for multiple comparisons with Bonferroni correction). Meanwhile, 23 of 28 different North American to North American comparisons resulted in significant F_{ST} values at $P \leq 0.05$, and when we corrected these for multiple comparisons, 18 comparisons remained significant. When pairwise comparisons from North American sites equal to or less than 1,200 km apart (the maximum distance separating Eurasian sites) were included, 13 of 21 comparisons were significant. Of the 40 North

American to Eurasian comparisons, 31 were significant at $P \leq 0.05$, and 17 of these remained significant after the correction for multiple comparisons. No relationship existed between geographical distance and pairwise genetic similarity measure (linearized F_{ST} , Slatkin 1993) in either range of the species (Fig. 2, North America: $Z = 65,614.38$, $r = -0.23$, $P = 0.184$; Eurasia: $Z = 17,382.73$, $r = 0.14$, $P = 0.699$).

The Structure analyses determined that the dataset was consistent with eight genetic clusters ($K = 8$ with 94.8% probability). Proportion membership of each individual to each of the eight clusters is shown in Fig. 3. Overall, most individuals did not assign strongly to a single genetic cluster, indicating a lack

Table 3 Pairwise F_{ST} values, by sampling location (see Table 1 for site abbreviations)

	USWY1	USWA1	USCO1	USMT1	USWA2	USCA2	USCO2	USAZ1	TR6	UA3	BG34	RUS15
USWA1	0.0131											
USCO1	0.0455	0.0959*										
USMT1	0.1171	0.1132*	0.2134*									
USWA2	0.0086	0.0076	0.1035*	0.0939*								
USCA2	0.1675*	0.1381*	0.2139*	0.1117*	0.0437							
USCO2	0.0682	0.0852*	0.1211*	0.1410*	0.0726*	0.1250*						
USAZ1	0.0736	0.0676*	0.0788*	0.0850*	0.0268	0.0909	0.0867*					
TR6	0.1397*	0.0678	0.0918*	0.0219	-0.0397	0.1954*	0.0274	0.0258				
UA3	0.0861	0.1064*	0.1833*	0.0920*	0.0579	0.0513	0.0829*	0.0633	-0.0253			
BG34	0.0927	0.0942*	0.1558*	0.1106*	0.0710	0.1206*	0.0947*	0.0727	0.0406	0.0820*		
RUS15	0.1006	0.0659	0.0872	0.0192	-0.0153	0.0890	0.0160	0.0258	-0.0830	0.0374	0.0172	
RUS5	0.0828	0.0889*	0.1400*	0.0782*	0.0306	0.0779*	0.0635*	0.0418	-0.0537	0.0591*	0.0662*	-0.0227

The upper left corner of the table shows NA \times NA comparisons, the lower left block shows NA \times EU comparisons, and the lower right shows EU \times EU comparisons. Bold values are significant at $P \leq 0.05$; starred values (*) are significant at $P \leq 0.05$ after Bonferroni correction for multiple comparisons

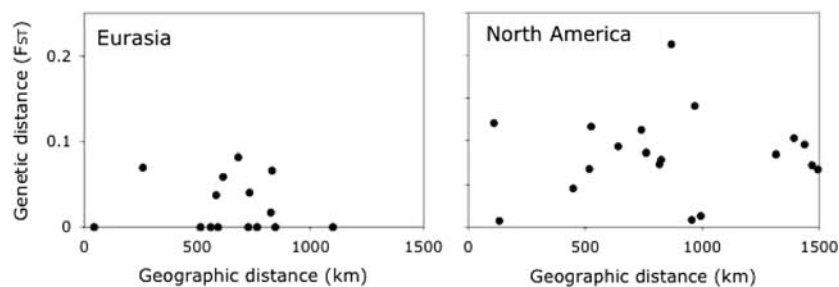


Fig. 2 Relationship between genetic and geographic distances in *Centaurea diffusa*. Pairwise F_{ST} is plotted against spatial distance in kilometers for all within-North America and within-

Eurasia sampling location comparisons. All F_{ST} values < 0 were changed to zero for plotting. Isolation by distance was not significant in either range (see text for details)

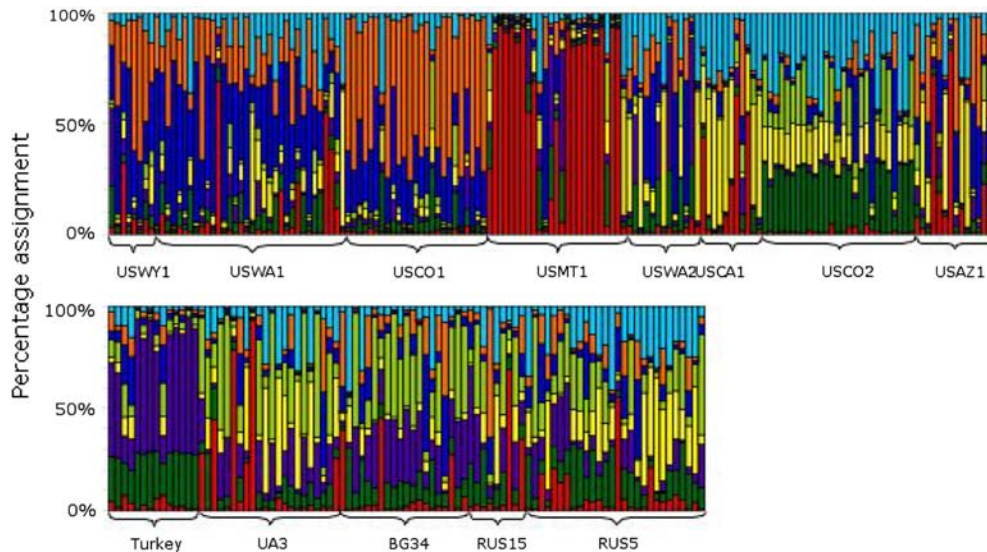


Fig. 3 Bar graph shows percent assignment of each individual to each genetic cluster (represented by different colors) when K (number of clusters) = 8. Each vertical bar represents one individual, and sampling location codes (see Table 1) are

indicated along the x -axis. North American individuals are shown at the top, and Eurasian individuals are shown at the bottom part of the figure

of genetic differentiation in the dataset. Most individuals from the Helena, MT site, however, did assign strongly to the cluster indicated by red.

Discussion

Departures from equilibrium

We observed several deviations from Hardy–Weinberg equilibrium in the dataset, all in the direction of heterozygote deficit. Seven of the deviant population by locus comparisons occurred at the same locus, 38CM22. This could be indicative of the presence of null alleles at this locus, or it could be due to selection. Null alleles result from mutations (usually insertions, deletions, or point mutations in a primer binding site) that prevent PCR amplification of the DNA fragments containing the microsatellites. These mutations mask the presence of heterozygotes, leading to a miscount of the actual number of heterozygotes present. However, if the locus is near a gene under selection a similar pattern could be seen. Three sampling sites deviated from Hardy–Weinberg equilibrium at several loci, indicating that one or more of the assumptions underlying Hardy–Weinberg equilibrium (no selection, no mutation, no migration,

no drift) are violated at these sites. One site was in the invasive range of the species and two sites were in the native range, an indication that native range *C. diffusa* are not necessarily stable or at equilibrium.

Genetic diversity similar in introduced and native ranges

Overall, allelic richness was similar between samples from the introduced range and samples from the native range of *C. diffusa*. Allelic diversity was slightly higher in the Eurasian samples, but not significantly so. The number of private alleles was higher in the sampled native range sites than in the introduced range sites. Even so, we detected many North American private alleles, suggesting that the European samples do not encompass all sources of the *C. diffusa* invasion into North America. There was no difference detected between the ranges in expected heterozygosity. Heterozygosity was higher, though not significantly so, in the invasive range, indicating that there has been no reduction in genetic diversity as a result of introduction. Sample size had a significant effect on heterozygosity and there was also a significant interaction between sample size and region, such that heterozygosity appeared to be more

sensitive to sample size in Eurasia than in North America. The average number of alleles per locus was dependent on sample size, but did not differ between the native range sampling locations and the invasive range sites, indicating no per-population reduction in genetic diversity. Other studies of invasive species also have found comparable levels of genetic diversity across both ranges, and generally attribute this pattern to a history of multiple introductions to the new range (Genton et al. 2005; Durka et al. 2005; Therriault et al. 2005). If multiple introductions occur, slight bottlenecks in each introduction will not necessarily reduce overall levels of genetic diversity, consistent with the pattern seen here. Interestingly, a study of *C. diffusa* cpDNA haplotypes found greater genetic diversity in the native range of *C. diffusa* than in the invasive range (Hufbauer and Sforza *in press*), demonstrating that marker choice can greatly influence the results of genetic diversity inquiries.

Population structure stronger in introduced range,
no pattern of isolation by distance

As discussed earlier, samples representing populations of organisms are likely to be genetically structured by geographic location if dispersal distances are smaller than the distance between sampling locations. Analysis of molecular variance showed that the majority of genetic variation in *C. diffusa* occurs within the sampling locations, which is not uncommon for an obligate outcrosser (Nyblom 2004). There was no significant continent level variance component, which is unsurprising given that Eurasia is the source for the North American invasion. Additionally, size homoplasy (different types of mutations leading to microsatellites of the same length) could cause differences between continents to be missed. When we compared variance partitioning between native and introduced ranges, however, we saw a surprising pattern. The invasive North American range had a higher degree of among-population variation than the native Eurasian range. Indeed, the sampled native Eurasian sites showed a small and only marginally significant level of population structure, indicating that they are probably close to panmictic. Meanwhile, over 10% of variation in North America was found between sampling

locations. Pairwise F_{ST} calculations also showed increased population genetic structuring in the introduced range of *C. diffusa*. We found that only about one third (3 of 10) of the Eurasian to Eurasian sampling site comparisons had a pairwise F_{ST} significantly different from zero. At the same time, nearly two thirds (18 of 28) of the North American to North American comparisons had significant pairwise F_{ST} , indicating a degree of population structure that was not observed in the native range. To ensure that this was not merely the effect of the larger spatial scale of North American sampling, we also examined the pairwise comparisons between North American sites less than 1,200 km apart (the maximum distance between Eurasian sampling sites). Here, 13 of 21 (over 60%) of comparisons were significant, indicating that the pattern of increased structure in the invasive range is not due to the greater distance between samples in North America. When we compared North American to Eurasian sampling sites, only 17 of the 40 pairwise F_{ST} values were significant, indicating less difference overall among continents than among sampling sites within North America. The results of the clustering software Structure are in agreement with AMOVA, pairwise F_{ST} , and isolation by distance results. The only site with most individuals strongly assigned to a single cluster was from Helena, MT (Fig. 3). None of the Eurasian sites showed consistent assignment of individuals to a single cluster, suggesting a high degree of gene flow. Indeed, gene flow estimates based on the numbers of private alleles (Slatkin 1985) were higher in Eurasia than in North America. This is an unexpected result; if population genetic structuring is not present in the native range, why is it present in the introduced range?

Isolation by distance is a pattern we might expect if populations are likely to be founded by or occasionally share genetic material with their closest neighbors. It is a more specific case of population structuring where geographic and genetic distances are positively correlated, and closely related individuals tend to be clustered spatially. Since there was negligible population structure in the native range of the species, we would not expect to see any significant isolation by distance, and we did not. We did observe population structure in the introduced range of the species, however, so the results of isolation by distance tests here are more interesting.

There was no relationship between geographic and genetic distance between sampling sites in North America (Fig. 2).

Centaurea diffusa, even in its native range, tends to be a roadside weed that flourishes in disturbed areas (Watson and Renney 1974). As a result, there may have never been much structuring in the native range, and the structure we see in the introduced range of the species may simply be a result of sampling effects and establishment of different genotypes in different areas after multiple introductions. Alternately, the native range *C. diffusa* might historically have been more genetically structured, but that structure may have been disrupted by human-mediated transport or interspecific hybridization with *C. stoebe stoebe* (diploid spotted knapweed) or other taxa in the genus *Centaurea*. The region sampled has been intensively disturbed via both agriculture and wars. Invasion biologists tend to think about native ranges of species as being more ‘natural,’ stable, and closer to equilibrium, but this may not always be the case. Few areas of the world remain unaffected by human development, and clearly, native range populations may be just as genetically disrupted as introduced range populations.

Multiple introductions are likely

The AMOVA and isolation by distance analyses showed that sampling sites within North America are genetically distinct, but sites that are close to one another spatially are not necessarily closer genetically. This pattern is consistent with multiple introductions of different genotypes from the native range into different areas of North America. In such a situation, population structure is the result of the chance establishment of different genotypes in different areas. This pattern could be reinforced with long-range ‘leapfrog’ dispersal, likely mediated by human transport of propagules. When populations are not founded by their nearest neighbors, a correlation between geographic and genetic distance will not develop.

Several relationships can be inferred from the Structure results. The cluster assignments of individuals from Bingen, WA, Afton, WY, and Fort Collins, CO are similar, suggesting that those populations are related. *C. diffusa* was first recorded in North

America from Bingen, WA, and the sampling site there shows mixed assignment. The two prominent clusters are indicated with blue and orange (Fig. 3). We cannot discern the probable native range origin for these individuals, as the genetic clusters are poorly represented in our samples from Eurasia, but the same clusters appear in the Afton, WY and Fort Collins, CO sampling sites, indicating that they may have been founded from propagules originating at this initial region of introduction. Interestingly, the Red Feather, CO site, which is less than 40 miles from Fort Collins, does not share assignment to either of these clusters. This is additional evidence for leapfrog dispersal. The Helena, MT site assigned most strongly to a single cluster, indicating a different introduction event for this population. None of the Eurasian locations sampled correspond to the Helena, MT cluster. Chloroplast DNA evidence also supports the hypothesis of multiple introductions in *C. diffusa* (Hufbauer and Sforza *in press*).

Given the extremely low degree of population structuring in the native range of this species and the evidence from private alleles that we did not sample some regions that contributed to this invasion, it is unsurprising that assignment tests failed to positively assign our North American sampling sites to probable Eurasian progenitors. If the native range were sampled more extensively, we might eventually be able to track down the origins of our North American invasion(s), but we might not. If near-panmixia is found across the Eurasian range then it may not be possible to pinpoint origins of invasive populations even with more samples.

Conclusion

Our results suggest that *C. diffusa* has been introduced multiple times into North America. This scenario is supported by the lack of a significant reduction in genetic diversity in the introduced range, and by assignment tests and *F*-statistics revealing strong population structure within North America. Population structure was found to be stronger in the species’ introduced range than in its native range. This higher invasive range structure may be a result of founder effects associated with separate introductions. The results are consistent with two or more introductions, however given the lack of population

structure within the native range, we are unable to pinpoint origins of those introductions. Additional samples from further east in Asia than we were able to include here might help us to draw more definitive conclusions regarding the specific origins of North American *C. diffusa*, as well as the evolutionary forces likely to be at work in this system.

Acknowledgments This work was funded by the USDA-NRI and the Colorado State University Agricultural Experiment Station. The authors thank Dr. William C. Black IV, Steven Rauth, Amy Blair, and Dr. Rebecca Kao for suggestions that greatly improved this manuscript, and Dr. Andrew Norton and Dale Woods for collecting samples.

References

- Amsellem L, Noyer JL, Le Bourgeois T, Hossaert-McKey M (2000) Comparison of genetic diversity in the invasive weed *Rubus alceifolius* Poir. (Rosaceae) in its native range and in areas of introduction, using amplified fragment length polymorphism (AFLP) markers. *Mol Ecol* 9:443–455
- Blossey B, Notzold R (1995) Evolution of increased competitive ability in invasive nonindigenous plants: a hypothesis. *J Ecol* 83:887–889
- Bohonak AJ (2002) IBD (Isolation By Distance): a program for analyses of isolation by distance. *J Hered* 93:153–154
- Bosssdorf O, Auge H, Lafuma L, Rogers WE, Siemann E, Prati D (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia* 144:1–11
- Brown AHD, Marshall DR (1981) Evolutionary changes accompanying colonization in plants. In: Scudder GEC, Reveal JL (eds) *Evolution today, proceedings of the 2nd international congress of systematic and evolutionary biology*. Hunt Institute for Botanical Documentation, Carnegie-Mellon University, Pittsburgh
- DeWalt SJ, Hamrick JL (2004) Genetic variation of introduced Hawaiian and native Costa Rican populations of an invasive tropical shrub, *Clidemia hirta* (Melastomataceae). *Am J Bot* 91:1155–1162
- Durka W, Bosssdorf O, Prati D, Auge H (2005) Molecular evidence for multiple introductions of garlic mustard (*Alliaria petiolata*, Brassicaceae) to North America. *Mol Ecol* 14:1697–1706
- Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proc Natl Acad Sci USA* 97(13):7043–7050
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Genton BJ, Shykoff A, Giraud T (2005) High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Mol Ecol* 14:4275–4285
- Goolsby JA, Makinson JR, Hartley DM, Zonneveld R, Wright AD (2004) Pre-release evaluation and host range testing of *Floracarus perrepae* (Eriophyidae) genotypes for biological control of Old World climbing fern. In: Cullen JM, Briese DT, Kriticos DJ, Lonsdale WM, Morin L, Scott JK (eds) *Proceedings of the XI international symposium on biological control of weeds*. CSIRO Entomology, Canberra, Australia
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (Version 2.9.3). <http://www.unil.ch/izea/software/fstat.html>. Cited 23 May 2007
- Grapputo A, Boman S, Lindstrom L, Lyytinen A, Mappes J (2005) The voyage of an invasive species across continents: genetic diversity of North American and European Colorado potato beetle populations. *Mol Ecol* 14:4207–4219
- Hufbauer RA, Sforza R (in press) Multiple introductions of invasive *Centaureas* inferred from cpDNA haplotypes. *Divers Distrib*
- Husband BC, Barrett SCH (1991) Colonization history and population genetic structure of *Eichornia paniculata* in Jamaica. *Heredity* 66:287–296
- Kliber A, Eckert CG (2005) Interaction between founder effect and selection during biological invasion in an aquatic plant. *Evolution* 59:1900–1913
- Kolbe JJ, Glor RE, Schettino LRG, Lara AC, Larson A, Losos JB (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431:177–181
- Lang RF, Richard RD, Parker PE, Wendel L (2000) Release and establishment of diffuse and spotted knapweed bio-control agents by USDA, APHIS, PPQ in the United States. *Pan-Pac Entomol* 76:197–218
- LeJeune KD, Seastedt TR (2001) *Centaurea* species: the forb that won the West. *Conserv Biol* 15:1568–1574
- Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends Ecol Evol* 20:136–142
- Marrs RA, Hufbauer RA, Bogdanowicz SJ, Sforza R (2006) Nine polymorphic microsatellite markers in *Centaurea stoebe* L. (subspecies *C. s. stoebe* and *C. s. micranthos* (S. G. Gmelin ex Gugler) Hayek) and *C. diffusa* Lam. (Asteraceae). *Mol Ecol Notes* 6:837–840
- Merila J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *J Evol Biol* 14:892–903
- Müller-Scharer H, Schroeder D (1993) The biological-control of *Centaurea* spp in North America—do insects solve the problem. *Pestic Sci* 37:343–353
- Neel JV (1973) Private genetic variants and frequency of mutation among South-American Indians. *Proc Natl Acad Sci USA* 70:3311–3315
- Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* 29:1–10
- Neuffer B, Hurka H (1999) Colonization history and introduction dynamics of *Capsella bursa-pastoris* (Brassicaceae) in North America: isozymes and quantitative traits. *Mol Ecol* 8:1667–1681

- Nyblom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol Ecol* 13:1143–1155
- Pauwels M, Saumitou-Laprade P, Holl AC, Petit D, Bonnin I (2005) Multiple origin of metalicolous populations of the pseudometallophyte *Arabidopsis halleri* (Brassicaceae) in central Europe: the cpDNA testimony. *Mol Ecol* 14:4403–4414
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Raymond M, Rousset F (1995) GENEPOP Version 1.2: population genetics software for exact tests and ecumenicism. *J Hered* 86:248–249
- Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* 55:1095–1103
- Roché BF Jr, Roché CF (1999) Diffuse knapweed. In: Sheley RL, Petroff JK (eds) *Biology and management of noxious rangeland weeds*. Oregon State University Press, Corvallis, USA
- Roderick GK, Navajas M (2003) Genotypes in novel environments: genetics and evolution in biological control. *Nat Rev Genet* 4:889–899
- Slatkin M (1985) Rare alleles as indicators of gene flow. *Evolution* 39:53–65
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47:264–279
- Therriault TW, Orlova MI, Docker MF, MacIsaac HJ, Heath DD (2005) Invasion genetics of a freshwater mussel (*Dreissena rostriformis bugensis*) in eastern Europe: high gene flow and multiple introductions. *Heredity* 95:16–23
- Thompson DJ, Stout DG (1991) Duration of the juvenile period in *C. diffusa* (*Centaurea diffusa*). *Can J Bot* 69:368–371
- Watson AK, Renney AJ (1974) The biology of Canadian weeds: 6. *Centaurea diffusa* and *C. maculosa*. *Can J Plant Sci* 54:687–701
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Williams DA, Overholt WA, Cuda JP, Hughes CR (2005) Chloroplast and microsatellite DNA diversities reveal the introduction history of Brazilian peppertree (*Schinus terebinthifolius*) in Florida. *Mol Ecol* 14:3643–3656
- Zenger KR, Richardson BJ, Vachot-Griffin. AM (2003) A rapid population expansion retains genetic diversity within European rabbits in Australia. *Mol Ecol* 12:789–794